



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### The spread of mosquito-borne viruses in modern times

**Citation for published version:**

Rossi, G, Karki, S, Smith, RL, Brown, WM & Ruiz, MOH 2018, 'The spread of mosquito-borne viruses in modern times: a spatio-temporal analysis of dengue and chikungunya', *Spatial and spatio-temporal epidemiology*, vol. 26, no. August 2018. <https://doi.org/10.1016/j.sste.2018.06.002>

**Digital Object Identifier (DOI):**

[10.1016/j.sste.2018.06.002](https://doi.org/10.1016/j.sste.2018.06.002)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Spatial and spatio-temporal epidemiology

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



1   **Title**

2   The spread of mosquito-borne viruses in modern times: a spatio-temporal analysis of dengue and  
3   chikungunya

4   **Authors**

5   Rossi, G.<sup>1\*</sup>, Karki, S.<sup>1</sup>, Smith, R.L.<sup>1</sup>, Brown, W.M.<sup>1</sup>, Ruiz, M. O.<sup>1</sup>

6   <sup>1</sup> Department of Pathobiology, College of Veterinary Medicine, University of Illinois at Urbana-  
7   Champaign, 2001 S Lincoln Ave, 61802 Urbana (IL), USA

8   \*Current address: Royal (Dick) School of Veterinary Medicine, University of Edinburgh, Easter Bush  
9   Campus, Midlothian EH25 9RG, UK

10  <sup>#</sup>Corresponding author: g.rossi@ed.ac.uk

11  **Abstract**

12  Since the 1970s, mosquito-borne pathogens have spread to previously disease-free areas, as well as  
13  causing increased illness in endemic areas. In particular, dengue and chikungunya viruses, transmitted  
14  primarily by *Aedes aegypti* and secondarily by *Aedes albopictus* mosquitoes, represent a threat for up to  
15  a third of the world's population, and are a growing public health concern.

16  In this study, we assess the spatial and temporal factors related to the occurrences of historic dengue  
17  and chikungunya outbreaks in 76 nations focused geographically on the Indian Ocean, with outbreak  
18  data from 1959 to 2009. First, we describe the historical spatial and temporal patterns of outbreaks of  
19  dengue and chikungunya in the focal nations. Second, we use a boosted regression tree approach to  
20  assess the statistical relationships of nations' concurrent outbreak occurrences and annual occurrences  
21  with their spatial proximity to prior infections and climatic and socio-economic characteristics.

22  We demonstrate that higher population density and shorter distances among nations with outbreaks are  
23  the dominant factors that characterize both dengue and chikungunya outbreaks. In conclusion, our  
24  analysis provides crucial insights, which can be applied to improve nations' surveillance and  
25  preparedness for future vector-borne disease epidemics.

26

27  **Keywords**

28  Vector-borne diseases; Dengue virus; Chikungunya virus; Indian Ocean; Boosted regression trees

## 29 1. Introduction

30 The spread of mosquito-borne pathogens to new areas has increased markedly since the 1970s  
31 along with an overall increase in cases of illnesses. Two diseases that saw a dramatic expansion during  
32 this period are those associated with the dengue and chikungunya viruses. Both viruses are transmitted  
33 by *Aedes aegypti* mosquitoes with secondary transmission by *Aedes albopictus*, and the two diseases  
34 share an intertwined history and ecology (Manore et al., 2014; Carey, 1971).

35 Dengue virus (DENV) cases have been increasing steadily for decades, as the area of endemic  
36 transmission has expanded. At the time of this writing, more than 1/3 of the world's population was at  
37 risk of illness from dengue virus (DENV), with recent estimates of 390 million human infections  
38 annually (Bhatt et al., 2013; Gurugama et al., 2010; Monath, 1994). DENV infections were first  
39 reported in 1780, and contributed to large concurrent epidemics in port cities of Asia, Africa and North  
40 America in the late 1700s - coinciding with an increase in global commerce (Gubler, 1998). After a  
41 more recent surge in cases during the years of World War II, cases of dengue declined thanks to  
42 intensive efforts to reduce the *Aedes aegypti* mosquitoes, only to expand dramatically after 1980  
43 (Murray et al., 2013). DENV has four distinct serotypes (DENV-1 – DENV-4), and some of the recent  
44 increase in cases has taken place as these serotypes expanded and mixed globally (Messina et al.,  
45 2014).

46 Chikungunya virus (CHIKV) was not identified until 1952 near the border of Tanzania and  
47 Mozambique (Lahariya and Pradhan, 2006), but it was likely circulating globally earlier than the  
48 1950s, with cases of illness sometimes being confused with those caused by DENV. Even today,  
49 reported cases without laboratory confirmation can lead to misclassification (Roth et al., 2014;  
50 Halstead, 1980; Carey, 1971). After the early 1970s, there was little evidence of chikungunya  
51 epidemics until a dramatic reemergence in the Indian Ocean region in 2005. During this period millions  
52 of cases were recorded within a short period of time, with peaks of 47,000 new cases in a single week

53 (Higgs, 2006), and a very high attack rate in some locales – including some 244,000 cases among the  
54 approximately 800,000 residents of the island of Reunion (Simon et al., 2008). This expansion  
55 coincided with a viral mutation of CHIKV that resulted in more efficient transmission by *Aedes*  
56 *albopictus* mosquitoes and an increase in virulence (Kucharz and Cebula-Byrska, 2012). In 2007 it  
57 reached Europe, when a temperate region in Italy reported locally transmitted CHIKV following  
58 introduction by travelers from the Indian Ocean (Chevillon et al., 2008; Simon et al., 2008). CHIKV  
59 cases were later observed in southern France, and in 2017, after a decade without local cases, Italy  
60 again saw invasion and subsequent local transmission of CHIKV (Amraoui and Failloux, 2016; Marano  
61 et al. 2017). CHIKV also spread to the Americas following the outbreak in the Indian Ocean, with  
62 nearly 1.7 million cases reported in this region between December 2013 and September 2015  
63 (Cassadou et al., 2014; Petersen and Powers, 2016; Roth et al., 2014).

64 Zika virus, also transmitted primarily by *Aedes aegypti*, more recently entered global awareness.  
65 It caused an extraordinary epidemic in Brazil, after first being reported in the spring of 2015, with  
66 thousands more cases in the Americas since then (Esposito and Fonseca, 2016; Faria et al. 2017). The  
67 most recent large Zika outbreaks followed its expansion to the Yap Islands in the Pacific Ocean in 2007  
68 and subsequent spread to other parts of Southeast Asia and the Pacific (Chang et al., 2016; Weaver et  
69 al., 2016).

70 Here, we focused on historical outbreaks of illness from DENV and CHIKV. Some general  
71 principles guide the processes that result in expansion of mosquito-borne pathogens, but the precise  
72 mechanisms behind the emergence, as well as the spatial and temporal dynamics of mosquito-borne  
73 diseases (MBDs) spread, can vary considerably (Wood et al., 2017). By comparing the patterns of two  
74 viruses with similar ecological niches, we expect to gain insight into general characteristics that  
75 promote their expansion. Zika has a similar transmission pattern as chikungunya and dengue, but it  
76 lacks the long historic record of global circulation, and we thus did not include it in this study. We did

77 not consider malaria, which is transmitted by different mosquito species and is not a viral disease.

78 MBDs are found primarily in tropical or sub-tropical regions, and areas where they have  
79 emerged and reemerge in recent years are typified by suitable vector habitat, large, diverse wildlife  
80 populations, and increasing human population density (Jones et al., 2008). On the other hand,  
81 socioeconomic factors such as lifestyle, poor infrastructure, and poor sanitation are also likely to  
82 facilitate their diffusion (Moreno-Madriñán and Turell, 2017). More recently, the movement of both  
83 mosquitoes and infected hosts has been enhanced by globalization and increased air traffic (Brown et  
84 al., 2012; Tatem et al., 2012). In addition, MBDs have the potential to increase with climate change and  
85 the accompanying changes in wet-dry periods (Donat et al., 2016; Naish et al., 2014; Patz et al., 1996;  
86 Tjaden et al., 2017). Other factors that can play a critical role for MBD transmission are the evolution  
87 of virulence and the increased resistance of mosquitoes to insecticides (Maciel-de-Freitas et al., 2014;  
88 Greenwood et al., 2008, Moncayo et al., 2004). Finally, emergence can be driven by poor housing  
89 conditions found in crowded urban and sub-urban settlements, lack of mosquito control, and invasion  
90 of vectors into new areas along with deforestation and development of new agricultural enterprises (Ali  
91 et al., 2017; Dash et al., 2013; Moreno-Madriñán and Turell, 2017; Petersen and Powers 2016; Schrag  
92 and Wiener, 1995).

93 We compared factors driving the historic emergence of dengue and chikungunya viruses after  
94 World War II. On the one hand, we had the expectation that similarities between the two would exist  
95 because of their similar transmission cycles. On the other hand, we recognize that the two viruses are in  
96 different families, with DENV belonging to the flaviviridae and CHIKV to the alphaviridae, and  
97 while their spatial distribution does have overlap, the temporal patterns of the two are distinct. The  
98 *Aedes albopictus* mosquitoes from multiple regions are quite competent in transmission of CHIKV, and  
99 generally are more competent in transmission of CHIKV than DENV. *Aedes albopictus* has a broader  
100 range across climatic conditions than *Aedes aegypti* (Brady et al. 2014; Turell et al., 1992). Now, it has

101 established itself globally, so the possibility of outbreaks in more temperate conditions, though still  
102 generally low, is higher for CHIKV (Kucharz and Cebula-Byrska, 2012; Bonizzoni et al., 2013; Manni  
103 et al., 2017). By comparing the exogenous factors associated with outbreaks of the two viruses across a  
104 significant time period, we expect to reveal general patterns of emergence that can help explain these  
105 two specific examples of expanding MBDs, and provide insights about the potential for expansion of  
106 other mosquito-borne viruses.

107         The objectives of this analysis are twofold. First, we described the spatial and temporal patterns  
108 of outbreaks of illness from DENV and CHIKV after the end of World War II in 76 nations centered on  
109 continents that border the Indian Ocean. Second, we have assessed the characteristics of these nations  
110 relative to their temporally concurrent outbreaks (or endemic status) of dengue and of chikungunya. By  
111 doing so, we provided a basis for comparison of the biological and social factors that might influence  
112 the spread of both the vector mosquitoes and the viruses, during a period of increasing cases from 1965  
113 to 2009.

## 114 **2. Materials and methods**

### 115 ***2.1. Dengue and chikungunya outbreaks dataset***

116         We developed an outbreak occurrence dataset that included an indication of outbreaks (or  
117 continuing endemic status) of dengue and chikungunya for each year from 1952 to 2009. For the sake  
118 of simplicity and ease of communication, we refer henceforth to the event of interest as an “outbreak”  
119 and use it to denote the observed presence of locally acquired CHIKV or DENV in a country in a year.  
120 Our definition did not consider the size of outbreaks, and reports of one to three cases were not  
121 included. A nation with hyper-endemic condition of ongoing annual dengue disease cases was  
122 considered in outbreak status from that point on. We also did not distinguish among dengue serotypes,  
123 because those data are very rare, especially in the older records.

124           The study region comprised 76 nations stretching from Africa to Australia, generally focused on  
125 the Indian Ocean (See Supplementary material: Appendix A and Figure 1). The status of each nation for  
126 each year was determined by a two-stage review of literature, supplemented by records from the  
127 regional offices of the World Health Organization and review of health data from national health  
128 ministries, when accessible. The first stage of the review was a search in the National Center for  
129 Biotechnology Information PubMed citations database on terms that showed papers about the history  
130 of the presence of chikungunya or dengue cases globally. This returned 63 documents. A second review  
131 focused specifically on searches that included the names (including prior names when those changed)  
132 of the 76 nations of interest and resulted in 84 additional papers. The review emphasized tables and  
133 maps that indicated places where dengue or chikungunya were reported. Finally, we consulted the  
134 World Health Organization regional data records for the African, Eastern Mediterranean, the South-  
135 East Asian and the Western Pacific regions and used these to both review and supplement the material  
136 from the literature. The outcome of interest indicated whether a given nation recorded an outbreak of  
137 illness in a given year or not. For the complete list of the evaluated documents and the datasets used,  
138 see Supplementary material: Appendix A.

## 139 ***2.2. Statistical Analysis***

140           We first evaluated the pattern of outbreaks by place, time, and the annual co-occurrence across  
141 nations using graphs and GIS mapping (Esri, ArcGIS). We then developed two groups of statistical  
142 models, with outbreaks from DENV and CHIKV analysed separately: 1) the co-occurrence of  
143 outbreaks in pairs of nations (co-occurrence models), and 2) the outbreaks relative to factors associated  
144 with each nation (nation-specific models). The statistical analyses focused on the years from 1965 to  
145 2009, when data on the selected covariates were available consistently for all of the nations under  
146 consideration. The co-occurrence model was designed to reveal the factors driving the spatio-temporal  
147 patterns of the two viruses through an assessment of characteristics between countries that had co-

148 occurrences of outbreaks of the viruses in a given year. To develop the co-occurrence models, we  
149 divided the binary yearly outbreaks time series into nine 5-year periods, from 1965 to 2009, and in each  
150 time period we calculated the number of years outbreaks co-occurred in each possible pair of countries  
151 during that 5-year period (from a minimum of 0, to a maximum of 5). The nation-specific models, on  
152 the other hand, aimed to determine the factors best able to predict the number of outbreaks of dengue  
153 and chikungunya across the observed region based on characteristics of the nations across time. For this  
154 analysis, we considered the yearly occurrence of outbreaks in each country during the entire  
155 observation period.

156         Boosted regression tree (BRT) methods were used to fit all models. The boosting technique uses  
157 a machine learning algorithm to produce a final prediction model that is an ensemble of individual  
158 regression trees in a stage wise fashion: the original data are fitted with a first regression tree, and then  
159 the residuals of that first model become the input data on which the second tree is fitted, and so on  
160 (Conley et al., 2014). One important feature of this type of model is that data are weighted repeatedly  
161 in each re-fitting on the previous tree residuals. In this way, the misclassified points in previous trees,  
162 have more weight than values that were classified correctly in the following fit. The learning rate  
163 controls the contribution of each tree to the final model (Conley et al., 2014). For the co-occurrences  
164 model we used a BRT model with a 10 fold cross-validation, with the BRT model parameters of  
165 learning rate, tree complexity, and bag fraction set as suggested by Elith et al. (2008). Boosted  
166 regression uses cross validation to minimize over-fitting by determining when adding additional trees  
167 no longer improves predictive performance, and selecting that optimum number of trees. Because the  
168 co-occurrences data were counts, we specified a Poisson type of model for the dependent variable, as  
169 done by Ashby et al. (2017). Each of the 5-year periods was treated as a separate model.

170         In the nation-specific model, the BRT approach was used as described above, but the form of  
171 the model was logistic regression, in which the probability that an outbreak of disease occurred in



172 nation  $i$  and year  $y$  (with corresponding covariates  $Z_{i,y}$ ) corresponded to  $P(\text{DEN}_{i,y}=1|Z_{i,y})$  or  
173  $P(\text{CHIK}_{i,y}=1|Z_{i,y})$  and was modelled with a logit function. To develop the nation-specific models, the  
174 dataset was divided in two parts: a first training part included data for 26 out of the 76 nations. The  
175 other 50 nations were used to test the model results. We repeated this procedure 50 times, in order to  
176 reduce the stochastic effect due to the nations' selection. The nation-specific models were evaluated  
177 against the observed data using point biserial correlation (Bahn and McGill, 2013), Test COR, between  
178 the observed values and the predicted probability ("polycor" package, R Core Team 2016), and the  
179 Bernoulli deviance (i.e. Test dev) that measured the residual deviance between the predicted values of  
180 the model and the observed values of the test data ('dismo' package, Hijmans et al., 2013). All BRT  
181 models were fitted in R using the 'gbm' and 'dismo' libraries (Hijmans et al., 2013; Ridgeway, 2013).

#### 182 2.2.1. Covariates for the co-occurrence models

183 For the co-occurrence models, all covariates measured factors that compared nation pairs (Table  
184 1). With the exception of the data for geographical and historical factors, which did not vary over time,  
185 we calculated the 5-year period averages for each of the other factors to fit the same time-step as the  
186 co-occurrence measures. Population counts and density, gross domestic product (GDP), and climatic  
187 data were all originally found at the year/country resolution.

188 First, we considered geographical and historical factors using data retrieved from the French  
189 research centre CEPII dataset (Meyer and Zignago, 2011). These factors were invariant through time  
190 including: the *geographical distance* between each pair (DIS) calculated as the distance between  
191 capital cities; *contiguity* (CON), a binary variable coded as 1 if two countries shared a border, 0  
192 otherwise; *common colonizer or formerly part of the same country* (CCO), a binary variable set to 1 if  
193 the two countries were under the rule of the same colonizing nation or had been part of the same  
194 country in the past, 0 otherwise; *common language* (CLA), a binary variable set to 1 if the two  
195 countries have a common language among the two most commonly spoken, 0 otherwise.

196 Second, we considered climatic factors. We first retrieved the yearly average temperature  
 197 (TEM) and precipitation (PRE) for each nation from the Climate Data API (World Bank). To compare  
 198 the countries with respect to these factors, we computed a climate similarity index *CLI*. This was  
 199 simply the Euclidean distance between countries in a space where TEM and PRE were the dimensions,  
 200 following Tatem et al. (2006). In order to properly balance the effect of the two, both TEM and PRE  
 201 were rescaled to a 0–1 range.

202 Third, we used the Gross Domestic Product as a measure of wealth of each nation to calculate  
 203 the difference between each pair of countries (DWE). Finally, we incorporated two demographic  
 204 variables: human population density (DEN) and migration flow (MIG). To compare the pairwise effect  
 205 of density, we used a gravity model variable (DEG) based on distance and density (Cauchemez et al.,  
 206 2014). Specifically the index between countries *i* and *j* was calculated as:

$$207 \quad DEG_{ij} = \frac{DEN_i \times DEN_j}{DIS_{ij}^2} \quad . \quad (1)$$

208 Net migration data were already in the form of bilateral flows (Abel 2013; Abel and Sander  
 209 2014), between the nation pairs. Because the original data for 1960-1999 (Abel 2013) were an estimate  
 210 of the net migratory flows for 10-year-periods between each pair of countries, we halved the 10-year  
 211 periods values between 1965 and 1989. We were able to keep the original flow values, estimated over  
 212 5-year periods from 1990 to 2009 (Abel and Sander 2014).

### 213 2.2.2 Covariates for the nation-specific models

214 The nation-specific models included eight covariates measured for each nation over time. First,  
 215 we collated the DEN, TEM, PRE, and GDP for each nation as described above, maintaining the  
 216 original yearly values. As a measure of international migration, we used the yearly total of incoming  
 217 and outgoing flows (MIG<sub>i</sub> and MIG<sub>o</sub>, respectively). To obtain those yearly values, we divided all nation  
 218 pairs' 10-year or 5-year flows by ten and five, respectively, and then summed all the incoming and

219 outgoing yearly migration values in which the 76 nations included in this study were listed both as an  
 220 origin and a destination.

221 In order to address the temporal autocorrelation of disease outbreaks we introduced two indices,  
 222 each computed separately for dengue and chikungunya. The first was a binary covariate that indicated  
 223 whether the nation was exposed to an outbreak in the previous year, *PRY* (equal to 1 for year *T* if an  
 224 outbreak was observed in year *T*-1, 0 otherwise). The second index was used to quantify the external  
 225 “force of infection” (ESI), based on other nations’ previous year outbreak status. For this index, we  
 226 followed Cauchemez et al. (2014) using a gravity model approach. In particular, for nation *i* and year *y*  
 227 the index was calculated as:

$$228 \quad ESI_{i,y} = \sum_{j \neq i}^T \left( \frac{V_{j,y-1} \times DEN_{j,y-1} \times DEN_{i,y-1}}{DIS_{ij}^2} \right),$$

229 (2)

230 in which  $V_{j,y-1}$ ,  $DEN_{j,y-1}$ , and  $DIS_{j,y-1}$  correspond to the outbreak status of the previous year in all nations  
 231  $j \in T \neq i$ , population density (#individuals per squared km), and between-country distance (in km; see  
 232 Table 1), respectively.

233 We also aimed to understand the particular importance of the two indices, *PRY* and *ESI*, relative  
 234 to the ability of the model to predict the occurrence of the dengue and chikungunya outbreaks for a  
 235 given year. Thus, we ran the following five models for each disease with the exclusion of *PRY* (models  
 236 2 and 4) and/or *ESI* (models 3 and 4) as described below:

- 237 1. The complete model (C) included all eight covariates (DEN, TEM, PRE, GDP, MIG<sub>i</sub>, MIG<sub>o</sub>,  
 238 PRY, ESI);
- 239 2. The second model (W) included seven covariates and excluded the external force of infection  
 240 (DEN, TEM, PRE, GDP, MIG<sub>i</sub>, MIG<sub>o</sub>, PRY);

3. The third model (E) excluded the nation's previous state (*PRY*) but included the external force of infection (DEN, TEM, PRE, GDP, MIG<sub>i</sub>, MIG<sub>o</sub>, ESI);
4. The fourth model (N) excluded both types of prior states, internal and external (DEN, TEM, PRE, GDP, MIG<sub>i</sub>, MIG<sub>o</sub>);
5. Finally, the fifth model consisted of a simplified (S) model, developed by starting from the complete model (C) and then dropping the non-influential covariates following Elith et al. (2008).

### 3. Results

Dengue outbreaks were recorded in 38 of the 76 nations under consideration during the study period from 1952 to 2009, and there were between 1 and 45 years of co-occurrence among nation pairs during that period. Outbreaks occurred in East and Southeast Asian nations throughout the entire period, but only Singapore and Thailand had outbreaks in all 45 years. South and Central Asian countries experienced dengue outbreaks consistently in the period from 1962 to 1972, then again starting in 1988 and continuing to the present (Figure 1). Sri Lanka had the most years of outbreaks for the South Asian region, for a total of 32 years. Only a few African nations had outbreaks, with only Nigeria, Djibouti, Somalia, and the Seychelles with four or more years of outbreaks.

Chikungunya outbreaks were recorded in 34 of the 76 nations during the period from 1952 to 2009, and there were from 1 to 5 years of co-occurrence during the 45-year study period. The five nations with 9 or more years of outbreaks recorded included Indonesia, Malaysia, Thailand, South Africa, and India, but African nations were represented in higher proportion than other regions and compared to dengue in most of the period (Figure 1). South Africa, Reunion, and Democratic Republic of the Congo were the three African nations with the highest number of years of outbreaks, with 10, 5,

264 and 5 years, respectively.

### 265 ***3.1. Co-occurrence models***

266 The dengue co-occurrence models were stable, and the percentage of the explained deviance  
267 obtained with the 10-fold cross validation was between 40% and 80%, with the highest explained  
268 deviance in models for the periods from 1975-1979 and from 1995-1999 (Figure 2). The most  
269 important covariate was the gravity index (DEG), whereby nations that were closer together  
270 geographically and had higher population density had more instances of dengue co-occurrence (Figure  
271 2). This result was consistent across all time periods, but the effect size declined over time. As  
272 expected, the partial dependencies plots (Supplementary material: Appendix B) showed that DEG  
273 generally had a positive effect on the number of co-occurrences. The geographical distance (DIS) and  
274 GDP difference (DWE) both were important in the dengue model. For DIS, the effect was negative at  
275 distances greater than 5,000 km, but was variable when closer than that. For DWE, the effect was  
276 positive for differences close to zero, but when differences were large; co-occurrence was variable.

277 The results of the co-occurrence models for chikungunya outbreaks indicated that the most  
278 important covariates were DIS, DEG and climatic distance (CLI). This model was not stable over time,  
279 however, with the explained CV deviance less than 40% in all but the final time period (2005 – 2009),  
280 and with the period 1990-1994 without any co-occurrences (Figure 3). The partial dependencies plots  
281 (Supplementary material: Appendix B) reflect this instability as well, showing very different patterns  
282 for each covariate, depending on the 5-year period. We thus developed eight annual models for a period  
283 restricted to 2002 to 2009. As with the first chikungunya model, the most important covariates were  
284 DEG, DIS and CLI, with DEG being especially important in the years with the most outbreak activity.  
285 As seen with the first chikungunya analysis, the stability of the analysis increased when the annual  
286 number of outbreak observations increased (Figure 4).

### 287 ***3.2. Nation specific models***

288 For all five versions of the nation specific models for dengue, no single model outperformed the  
289 others if we take into account only the area under the receiver operating characteristic curve (AUC) of  
290 the cross validation (Figure 5, A). However, by considering the AUC that resulted from the independent  
291 test, models W and N were notably weaker. These two weak models did not include the external force  
292 of infection from other countries (ESI), while the weakest model, N, had neither ESI nor PRY (Figure  
293 5, B). In agreement with this result, the same two models produced the lowest correlated result relative  
294 to the observed dengue outbreaks occurrence values (Figure 6, A). Among all models, the complete  
295 model produced the best results.

296 The chikungunya nation-specific models produced similar results to the dengue models (Figure  
297 5, C and D). However, the prediction power was lower (Figure 6, B). As for DEN, the complete model  
298 was the best for the chikungunya analysis, but this result was not as clear as for dengue. For both of the  
299 complete models, the ESI index, measuring the external force of infection, had the highest relative  
300 contribution (Figure 7, A and B), while the other covariates played a relatively minor role.

301 The simplified (S) version of the nation-specific models automatically withdrew the less  
302 influencing covariates from the complete model. In the case of dengue, PRY was withdrawn 38 of 50  
303 times, followed by TEM (16), DEN (15), MIG<sub>i</sub> (11), GDP (10), MIG<sub>o</sub> (3), and PRE (2). ESI was never  
304 withdrawn. Similarly, in the S models for chikungunya, ESI was never withdrawn, while PRY was  
305 withdrawn 40 times, followed by TEM (25), GDP (23), DEN (17), MIG<sub>o</sub> (15), MIG<sub>i</sub> (11), and PRE (3).  
306 Those numbers were reflected in the relative contribution provided by each covariate in the complete  
307 model (Figure 7).

308

#### 309 4. Discussion and Conclusions

310 The main objectives of this work were to: (i) show how dengue and chikungunya spread across

311 the Indian Ocean region during the time period following World War II, and (ii) provide an estimate of  
312 which factors might have facilitated this spread. From the analysis of outbreak co-occurrences, we  
313 learned that dengue and chikungunya outbreaks were more likely to take place between pairs of nations  
314 that were relatively close to each other and had high population densities (DEG). This factor, based on  
315 the gravity model concept, had considerably more importance than simple contiguity. Considering  
316 chikungunya during the years from 2002 to 2009, DIS had higher relative importance than DEG during  
317 the years when fewer nations reported outbreaks, but DEG and DIS had about the same level of  
318 importance during years when more nations had outbreaks. This indicates that in the years when  
319 chikungunya was less widespread, proximity to another nation with an outbreak was a risk factor, but  
320 during years with higher levels of outbreaks, having both high population density and being nearer to  
321 other nations with high population density increased risk.

322         The external force of infection index (ESI), used in the nation-specific models, was a similar  
323 measure to DEG, but it also included the effect of the number of proximate nations with outbreaks in  
324 the prior year. Compared to DEG, ESI was much more clearly important relative to the other variables  
325 considered for both dengue and chikungunya annual outbreaks, and was even more important than the  
326 information about an outbreak within the same nation in the previous year (PRY). Because this index  
327 was built with information about both population density and distance, it might act as a proxy for non-  
328 observed variables, such as the number of travellers or shipped goods. In fact, while travelling can  
329 move infectious individuals and spread pathogens to favourable landscapes, shipping can move both  
330 infected people and vectors residing within containers, on ships, or in other transport equipment (Tatem  
331 et al., 2006). Given the absence of these types of data, we also used historical/anthropological  
332 variables, such as the common language, and common colonizer as covariates. The rationale behind  
333 this choice was that countries with languages in common or that were part of the same nation in the  
334 past might have more connections, or that this would be a facilitation for the movement of people and

335 goods. However, these variables did not prove important for any of our models. A more direct measure  
336 of movement of goods would have improved our ability to assess this, but our analysis went back to the  
337 1960s, and unfortunately, reliable, comprehensive datasets on flight connections, passenger numbers,  
338 shipping volumes and boat routes were not available for the entire study period. These types of data  
339 may be a further step for analysis of the more recent years, and could add more depth and precision in  
340 the analysis of that time period.

341 Our regression analyses spanned 45 years, from 1965 to 2009. Data on the subsequent years (to  
342 2011) were available for the presence/absence of dengue and chikungunya, however not all covariates  
343 were available (migration) or available in the same format (temperature and precipitation). Thus, we  
344 elected not to run different models at different points in time, as the comparison of the same factors  
345 through time was one of the key features of this work. For the same reason, we did not consider such  
346 data as the number of flying passengers, which can be found online for recent periods only. The choice  
347 of considering 5-years long periods in the co-occurrence model was very important for the analysis,  
348 because we minimized the time periods in which it was not possible to run the models due to no  
349 outbreak co-occurrences, as would have happened by considering a year-to-year time series. As  
350 indicated by the results, when the number of co-occurrences was low, the analysis lost power.

351 The precise definition of an outbreak year required subjective judgement, while the variability  
352 of surveillance consistency and the changing contemporary interest by the research community would  
353 have influenced the amount of information available for our review. By focusing systematically on only  
354 the 76 nations in question and having three members of the team come to conclusions both  
355 independently and through discussion, we reduced this problem. Moreover, given the above described  
356 issues, by considering the binary time series per country (presence/absence of dengue and  
357 chikungunya), our data gain temporal and spatial homogeneity, as the case reporting rates were very  
358 different at different points in time, and in different countries characterized by widely different



359 surveillance practices and level of research effort. On the one hand, this assumption might lead to bias  
360 because we neglected the number of cases, but on the other hand we gained in consistency across time,  
361 and that was a fundamental point of the analysis.

362 The WHO DengueNet was originally considered as a primary source of information for dengue  
363 outbreaks, but while it is a valuable resource to determine the status of dengue in a nation,  
364 completeness is inconsistent globally (Ruberto et al., 2015). Endemic status was another difficulty.  
365 While it is not difficult to identify a clear outbreak, such as chikungunya in Reunion Island in 2006 or  
366 Kolkata in 1963, identifying endemic dengue in light of both undercounting of cases and neglect of  
367 data can be problematic. India provides an important example: from the late 1980's on, our assessment  
368 was that India had dengue outbreak status every year, even though the years from 1997 to 2000 had  
369 relatively low numbers of reported cases compared to the years before and after (Chakravarti et al.,  
370 2012). We made this assessment because the total reported cases during those low years were still in  
371 the high hundreds of cases and the number of reported cases was most likely consistently under  
372 reported (Bhatt et al., 2013). This also illustrates a limitation on the designation of the “nation” as the  
373 observational unit. India’s outbreaks took place over a dispersed population, and variability in such a  
374 large, ecologically diverse and highly populated nation could be as important as between smaller  
375 nations, or even more. Another issue we had regarded but were not able to account for were the four  
376 different serotypes of DENV: only more recently have government reports and published literature  
377 included these consistently, but most often our sources did not report the serotype involved in each  
378 outbreak. Thus, we could not develop individual outbreak data or analyse the four serotypes separately.

379 Dengue infections have continued to increase globally since the end year of our analysis in  
380 2009, and today, the WHO considers at least 100 countries globally to have endemic transmission  
381 (WHO 2017). Chikungunya has likewise expanded since 2009 (Petersen and Powers 2016). Our  
382 analysis approach was appropriate for conditions of newly emerging or intermittent outbreaks, but

383 more detailed data are needed to determine the dynamics of the spread of dengue in recent years. In the  
384 absence of more complete and refined surveillance data on vector-borne diseases, the processes that  
385 lead to expansion of MBDs will remain incomplete.

386         The development of the covariates also required decisions and assumptions. For example, we  
387 built the two indexes (DEG and ESI) using the population density instead of the rough number of  
388 individuals. Our reasoning was twofold. First, the nations of interest had a wide population and area  
389 size range, so the number of individuals might not be very representative of actual crowding. Second,  
390 we detected a higher correlation between population and migration flows ( $0.67, p < 10^{-3}$ ), than between  
391 density and migration flows ( $0.23, p < 10^{-3}$ ). Thus, we chose the latter, population density, in order to  
392 reduce potential multicollinearity effects. For similar reasons, we chose the distance between capitals  
393 instead of the distance between the centre of each country, because the capital city might be more  
394 representative of the population distribution in heterogeneous countries. On the other hand, the two  
395 distances were very similar (correlation  $0.99, p < 10^{-2}$ ), and early exploratory models run with both  
396 showed no difference. The choice of using boosted regression tree technique was very important as  
397 well. Thanks to this technique, we could use different types of covariates with no need to transform  
398 them. Also, the interpretation of BRT results is generally very intuitive, and models include the  
399 potential interactions of covariates, without having explicitly to account for them, such as in classical  
400 logistic regression. Finally, machine learning techniques such as BRT can be easily re-applied to  
401 similar datasets, and their use is expanding in the epidemiological context.

402         In conclusion, our results highlighted trends found in both dengue and chikungunya outbreaks  
403 in the Indian Ocean region. In particular, the density of populations and less distance between  
404 outbreaks were dominant factors in the occurrence of those diseases, as found generally for emerging  
405 infectious diseases (Jones et al., 2008). We also found that outbreaks in neighbouring countries  
406 provided a good prediction of local outbreaks for both diseases. The relative importance of these

407 factors may reflect the higher probability of both vectors or infected people moving between countries  
408 as well the propensity for nearby nations to have similar climatic conditions. Even with increased  
409 global movement of people and goods, the ideas behind the gravity indexes were clearly important. In a  
410 similar vein, Cauchemez et al. (2014) found that distance between countries was a better predictor of  
411 chikungunya spread in the Caribbean than air transportation. Our analysis also highlighted how the  
412 boosted regression trees (BRT) approach can be used for both paired variables in the co-occurrence  
413 model as well as for the prediction model by nations.

414         The incorporation of data from available sources and further analyses in this direction will help  
415 the scientific community in reaching a comprehensive picture about the global spread of vector-borne  
416 diseases. This will help Ministries of Health, the WHO, and other agencies to more effectively allocate  
417 resources (i.e. financial, research, workforce) to prioritize those situations where conditions make  
418 outbreaks most likely in order to limit the diffusion of these epidemics.

419

## 420 **Acknowledgements**

421 The authors thank Stephanie Mundis who provided valuable help in carrying out the first review of the  
422 papers.

423 **References:**

- 424 1. Abel, G. J. 2013. Estimating global migration flow tables using place of birth data.  
425 Demographic Research 28: 505-546.
- 426 2. Abel, G. J., and N. Sander. 2014. Quantifying Global International Migration Flows. Science  
427 343: 1520-1522.
- 428 3. Ali, S., O. Gugliemini, S. Harber, A. Harrison, L. Houle, J. Ivory, S. Kersten, R. Khan, J. Kim,  
429 C. LeBoa, E. Nez-Whitfield, J. O'Marr, E. Rothenberg, R. M. Segnitz, S. Sila, A. Verwillow, M.  
430 Vogt, A. Yang, and E. A. Mordecai. 2017. Environmental and Social Change Drive the  
431 Explosive Emergence of Zika Virus in the Americas. PLoS neglected tropical diseases 11:  
432 e0005135.
- 433 4. Amraoui F., A. B. Failloux. 2016. Chikungunya: an unexpected emergence in Europe. Current  
434 Opinions in Virology 21:146–150. doi:10.1016/j.coviro.2016.09.014.
- 435 5. Ashby, J., M. J. Moreno-Madriñán, C. T. Yiannoutsos, and A. Stanforth. 2017. Niche Modeling  
436 of Dengue Fever Using Remotely Sensed Environmental Factors and Boosted Regression Trees.  
437 *Remote Sensing* 9: 328. doi: 10.3390/rs9040328
- 438 6. Bahn V, B. J. McGill. 2013. Testing the predictive performance of distribution models. *Oikos*  
439 122:321–331. doi:10.1111/j.1600-0706.2012.00299.x.
- 440 7. Bastian, M., S. Heymann, M. Jacomy. 2009. Gephi: an open source software for exploring and  
441 manipulating networks. In: International AAAI Conference on Weblogs and Social Media, San  
442 Jose, California. ISWSM.
- 443 8. Bhatt, S., P. W. Gething, O. J. Brady, J. P. Messina, A. W. Farlow, C. L. Moyes, J. M. Drake, J.  
444 S. Brownstein, A. G. Hoen, O. Sankoh, M. F. Myers, D. B. George, T. Jaenisch, G. R. Wint, C.  
445 P. Simmons, T. W. Scott, J. J. Farrar, and S. I. Hay. 2013. The global distribution and burden of  
446 dengue. *Nature* 496: 504-507.
- 447 9. Bonizzoni, M., G. Gasperi, X. Chen, and A. A. James. 2013. The invasive mosquito species  
448 *Aedes albopictus*: current knowledge and future perspectives. *Trends in parasitology* 29: 460-  
449 468.
- 450 10. Brady, O. J., N. Golding, D. M. Pigott, M. U. Kraemer, J. P. Messina, R. C. Reiner, Jr., T. W.  
451 Scott, D. L. Smith, P. W. Gething, and S. I. Hay. 2014. Global temperature constraints on *Aedes*  
452 *aegypti* and *Ae. albopictus* persistence and competence for dengue virus transmission. *Parasites*  
453 *& vectors* 7: 338.
- 454 11. Brown, E. B., A. Adkin, A. R. Fooks, B. Stephenson, J. M. Medlock, and E. L. Snary. 2012.

- Assessing the risks of West Nile virus-infected mosquitoes from transatlantic aircraft: implications for disease emergence in the United Kingdom. *Vector borne and zoonotic diseases* 12: 310-320.
12. Carey, D. E. 1971. Chikungunya and dengue: a case of mistaken identity? *Journal of the History of Medicine* 26: 243-262.
13. Cassadou, S., S. Boucau, M. Petit-Sinturel, P. Huc, I. Leparç-Goffart, and M. Ledrans. 2014. Emergence of chikungunya fever on the French side of Saint Martin island, October to December 2013. *Euro surveillance: bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* 19.
14. Cauchemez, S., M. Ledrans, C. Poletto, P. Quenel, H. de Valk, V. Colizza, and P. Y. Boelle. 2014. Local and regional spread of chikungunya fever in the Americas. *Eurosurveillance* 19: 15-23.
15. Chakravarti, A., R. Arora, and C. Luxemburger. 2012. Fifty years of dengue in India. *Trans.R.Soc.Trop.Med.Hyg.* 106: 273-282.
16. Chang, C., K. Ortiz, A. Ansari, and M. E. Gershwin. 2016. The Zika outbreak of the 21st century. *Journal of autoimmunity* 68: 1-13.
17. Chevillon, C., L. Briant, F. Renaud, and C. Devaux. 2008. The Chikungunya threat: an ecological and evolutionary perspective. *Trends in microbiology* 16: 80-88.
18. Conley, A. K., D. O. Fuller, N. Haddad, A. N. Hassan, A. M. Gad, and J. C. Beier. 2014. Modeling the distribution of the West Nile and Rift Valley Fever vector *Culex pipiens* in arid and semi-arid regions of the Middle East and North Africa. *Parasites & vectors* 7.
19. Dash, A. P., R. Bhatia, T. Sunyoto, and D. T. Mourya. 2013. Emerging and re-emerging arboviral diseases in Southeast Asia. *Journal of vector borne diseases* 50: 77-84.
20. Donat, M. G., A. L. Lowry, L. V. Alexander, P. A. O'Gorman, and N. Maher. 2016. More extreme precipitation in the world's dry and wet regions. *Nature Climate Change* 6: 508-+.
21. Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77: 802-813.
22. Esposito D. L. A., B. A. L. Da Fonseca. 2016. Zika and chikungunya infections in Brazil: Reviewing the epidemic and treatment options. *Revista da Sociedade Brasileira de Medicina Tropical* 49:535–536. doi:10.1590/0037-8682-0428-2016.
23. Faria N. R., J. Quick, I. M. Claro, J. Thézé, J. G. de Jesus, M. Giovanetti, et al. 2017. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* 546:406.

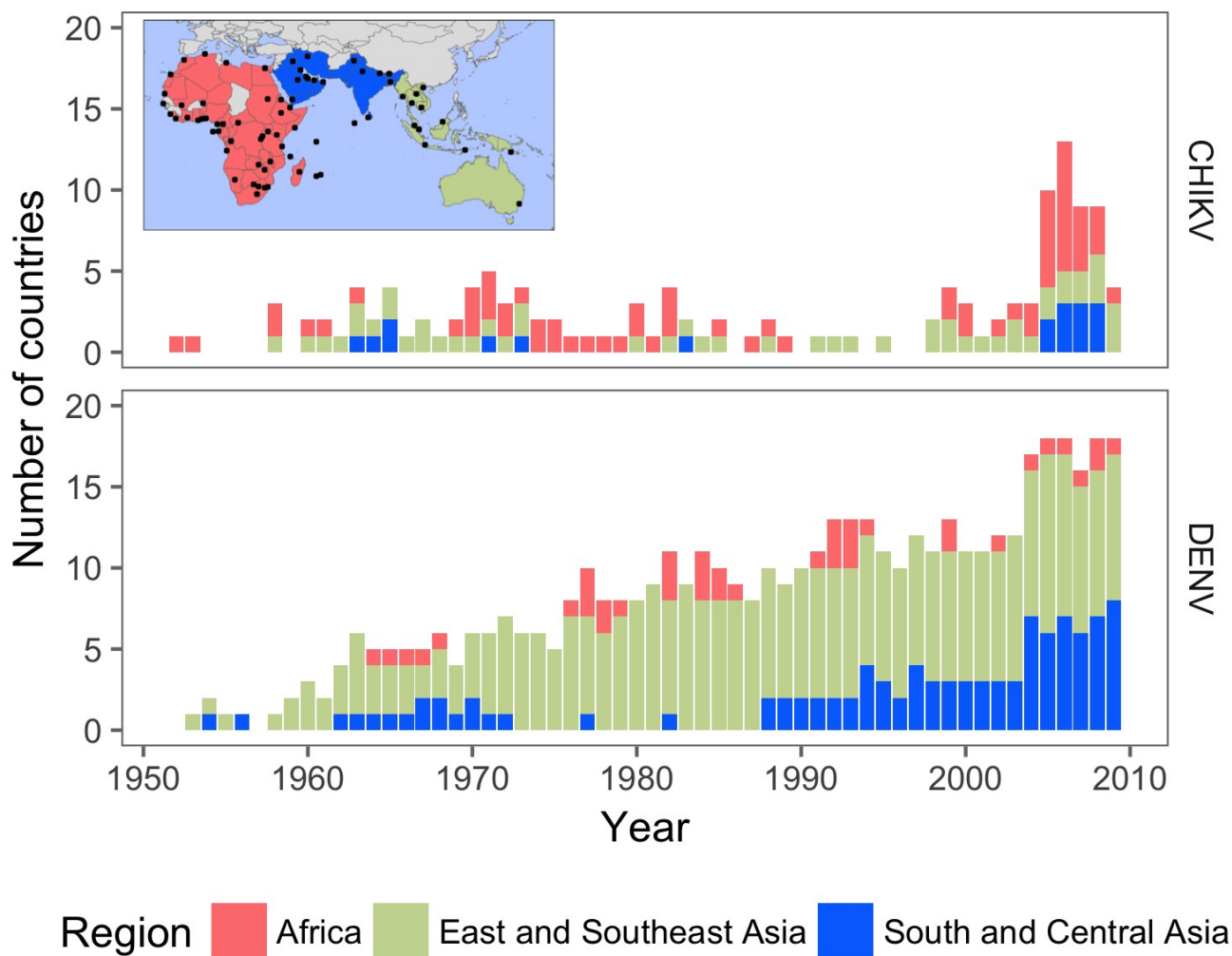
24. Greenwood, B. M., D. A. Fidock, D. E. Kyle, S. H. Kappe, P. L. Alonso, F. H. Collins, and P. E. Duffy. 2008. Malaria: progress, perils, and prospects for eradication. *The Journal of clinical investigation* 118: 1266-1276.
25. Gubler, D. J. 1998. Dengue and dengue hemorrhagic fever. *Clinical microbiology reviews* 11: 480-496.
26. Gurugama, P., P. Garg, J. Perera, A. Wijewickrama, and S. L. Seneviratne. 2010. Dengue viral infections. *Indian journal of dermatology* 55: 68-78.
27. Halstead, S. B. 1980. Dengue haemorrhagic fever--a public health problem and a field for research. *Bulletin of the World Health Organization* 58: 1-21.
28. Higgs, S. 2006. The 2005-2006 Chikungunya epidemic in the Indian Ocean. *Vector borne and zoonotic diseases* 6: 115-116.
29. Hijmans, R., S. Phillips, J. Leathwick, and J. Elith. 2014. Package 'dismo'. 2011.
30. Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak. 2008. Global trends in emerging infectious diseases. *Nature* 451: 990-993.
31. Kucharz, E. J., and I. Cebula-Byrska. 2012. Chikungunya fever. *European journal of internal medicine* 23: 325-329.
32. Lahariya, C., and S. K. Pradhan. 2006. Emergence of chikungunya virus in Indian subcontinent after 32 years: A review. *Journal of vector borne diseases* 43: 151-160.
33. Maciel-de-Freitas, R., F. C. Avendanho, R. Santos, G. Sylvestre, S. C. Araujo, J. B. Lima, A. J. Martins, G. E. Coelho, and D. Valle. 2014. Undesirable consequences of insecticide resistance following *Aedes aegypti* control activities due to a dengue outbreak. *PloS one* 9: e92424.
34. Mackenzie, J. S., and D. T. Williams. 2009. The zoonotic flaviviruses of southern, south-eastern and eastern Asia, and Australasia: the potential for emergent viruses. *Zoonoses and public health* 56: 338-356.
35. Manni, M., C. R. Guglielmino, F. Scolari, A. Vega-Rua, A. B. Failloux, P. Somboon, A. Lisa, G. Savini, M. Bonizzoni, L. M. Gomulski, A. R. Malacrida, and G. Gasperi. 2017. Genetic evidence for a worldwide chaotic dispersion pattern of the arbovirus vector, *Aedes albopictus*. *PLoS Neglected Tropical Diseases* 11: e0005332.
36. Marano, G., S. Pupella, I. Pati, F. Masiello, M. Franchini, S. Vaglio, C. Velati, and G. M. Liunbruno. 2017. Ten years since the last Chikungunya virus outbreak in Italy: history repeats itself. *Blood transfusion = Trasfusione del sangue* 15: 489-490.
37. Mayer, T., and S. Zignago. 2011. Notes on CEPII's distances measures: The GeoDist database.
38. Messina, J. P., O. J. Brady, T. W. Scott, C. Zou, D. M. Pigott, K. A. Duda, S. Bhatt, L.

- Katzelnick, R. E. Howes, K. E. Battle, C. P. Simmons, and S. I. Hay. 2014. Global spread of dengue virus types: mapping the 70 year history. *Trends in microbiology* 22: 138-146.
39. Monath, T. P. 1994. Dengue: the risk to developed and developing countries. *Proceedings of the National Academy of Sciences of the United States of America* 91: 2395-2400.
40. Moncayo, A. C., Z. Fernandez, D. Ortiz, M. Diallo, A. Sall, S. Hartman, C. T. Davis, L. Coffey, C. C. Mathiot, R. B. Tesh, and S. C. Weaver. 2004. Dengue emergence and adaptation to peridomestic mosquitoes. *Emerging infectious diseases* 10: 1790-1796.
41. Moreno-Madriñán, M. J., M. Turell. 2017. Factors of concern regarding Zika and other *Aedes aegypti*-Transmitted viruses in the United States. *Journal of Medical Entomology* 54: 251-257. doi: 10.1093/jme/tjw212.
42. Naish S., P. Dale, J. S. Mackenzie, J. McBride, K. Mengersen, S. Tong. 2014. Climate change and dengue: a critical and systematic review of quantitative modelling approaches. *BMC Infectious Diseases* 14:167. doi:10.1186/1471-2334-14-167.
43. Patz, J. A., P. R. Epstein, T. A. Burke, and J. M. Balbus. 1996. Global climate change and emerging infectious diseases. *JAMA: the journal of the American Medical Association* 275: 217-223.
44. Petersen, L. R., and A. M. Powers. 2016. Chikungunya: epidemiology. *F1000Research* 5.
45. Ridgeway, G. 2007. Generalized Boosted Models: A guide to the gbm package. Update 1: 2007.
46. Roth, A., A. Mercier, C. Lepers, D. Hoy, S. Duituturaga, E. Benyon, L. Guillaumot, and Y. Souares. 2014. Concurrent outbreaks of dengue, chikungunya and Zika virus infections - an unprecedented epidemic wave of mosquito-borne viruses in the Pacific 2012-2014. *Euro surveillance: bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* 19.
47. Ruberto, I., E. Marques, D. S. Burke, and W. G. Van Panhuis. 2015. The availability and consistency of dengue surveillance data provided online by the World Health Organization. *PLoS neglected tropical diseases* 9: e0003511.
48. Schrag, S. J., and P. Wiener. 1995. Emerging infectious disease: what are the relative roles of ecology and evolution? *Trends in ecology & evolution* 10: 319-324.
49. Simon, F., H. Savini, and P. Parola. 2008. Chikungunya: a paradigm of emergence and globalization of vector-borne diseases. *The Medical clinics of North America* 92: 1323-1343, ix.
50. Tatem, A. J., S. I. Hay, and D. J. Rogers. 2006. Global traffic and disease vector dispersal. *Proceedings of the National Academy of Sciences of the United States of America* 103: 6242-

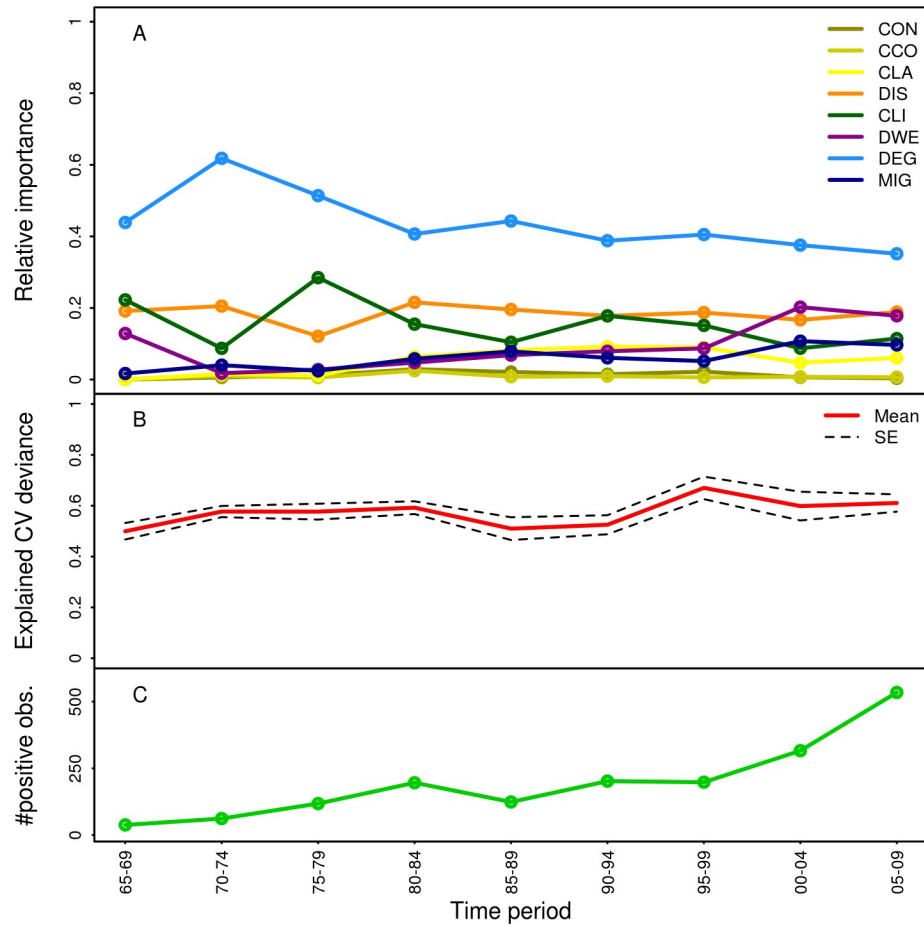
6247.

51. Tatem, A. J., Z. Huang, A. Das, Q. Qi, J. Roth, and Y. Qiu. 2012. Air travel and vector-borne disease movement. *Parasitology* 139: 1816-1830.
52. Tjaden N. B., J. E. Suk, D. Fischer, S. M. Thomas, C. Beierkuhnlein, J. C. Semenza. 2017. Modelling the effects of global climate change on Chikungunya transmission in the 21st century. *Scientific Reports* 7:3813. doi:10.1038/s41598-017-03566-3.
53. Turell, M. J., J. R. Beaman, and R. F. Tammariello. 1992. Susceptibility of selected strains of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) to chikungunya virus. *Journal Medical Entomology* 29: 49–53. doi: 10.1093/jmedent/29.1.49
54. Weaver, S. C. 2014. Arrival of chikungunya virus in the new world: prospects for spread and impact on public health. *PLoS neglected tropical diseases* 8: e2921.
55. Weaver, S. C., F. Costa, M. A. Garcia-Blanco, A. I. Ko, G. S. Ribeiro, G. Saade, P. Y. Shi, and N. Vasilakis. 2016. Zika virus: History, emergence, biology, and prospects for control. *Antiviral research* 130: 69-80.
56. Wood, C. L., A. McInturff, H. S. Young, D. Kim, and K. D. Lafferty. 2017. Human infectious disease burdens decrease with urbanization but not with biodiversity. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 372.
57. World Health, O. 2017. Dengue and severe dengue: Fact sheet. *In* WHO [ed.], <http://www.who.int/mediacentre/factsheets/fs117/en/>.

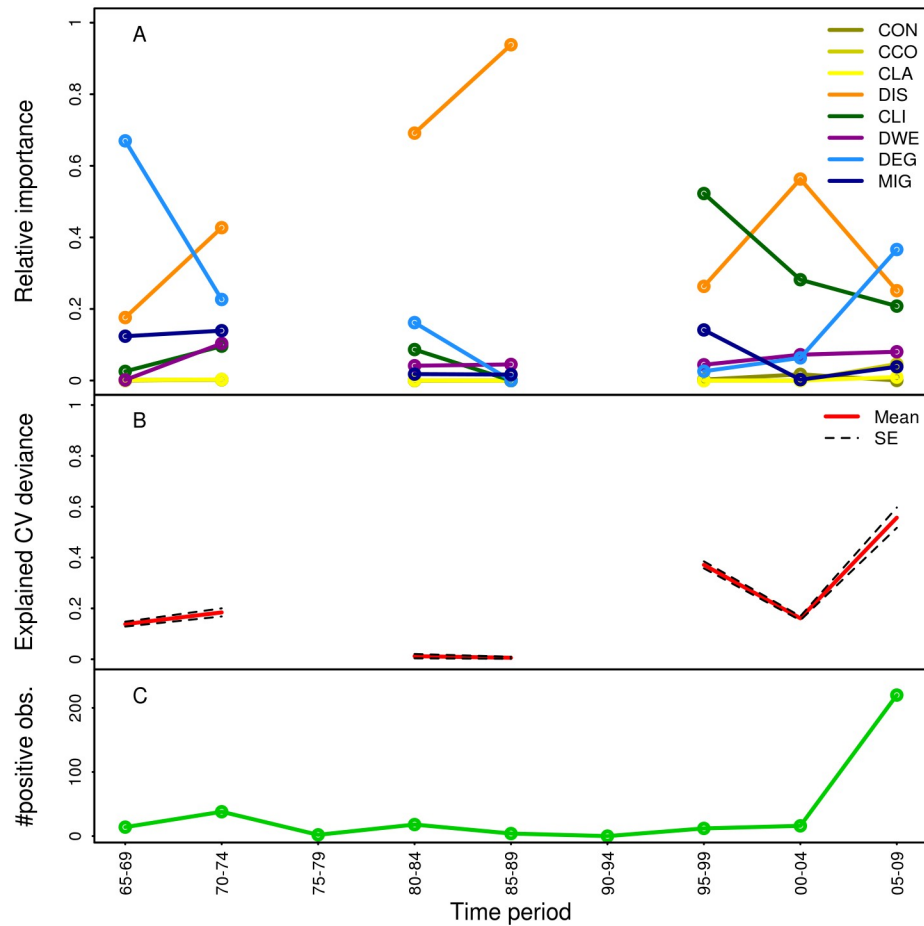




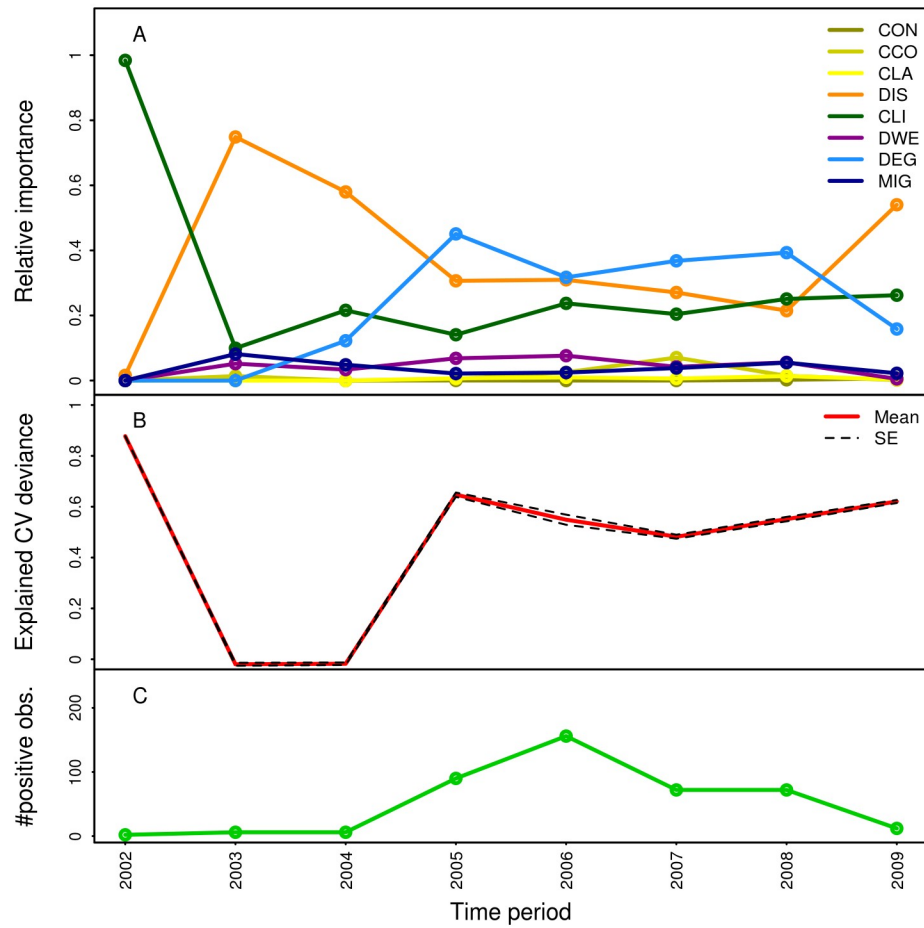
575 **Figure 1:** The 76 nations included in the study, the capital cities used for distance metrics (map in the  
576 top-left box) and the number of nations with an outbreak of chikungunya (top panel) and dengue  
577 (bottom panel) for the years from 1952 to 2009. The green portion of the bars are nations in Oceania,  
578 and in East and Southeast Asia, blue bars are in South and Central Asia and red bars are in Africa.



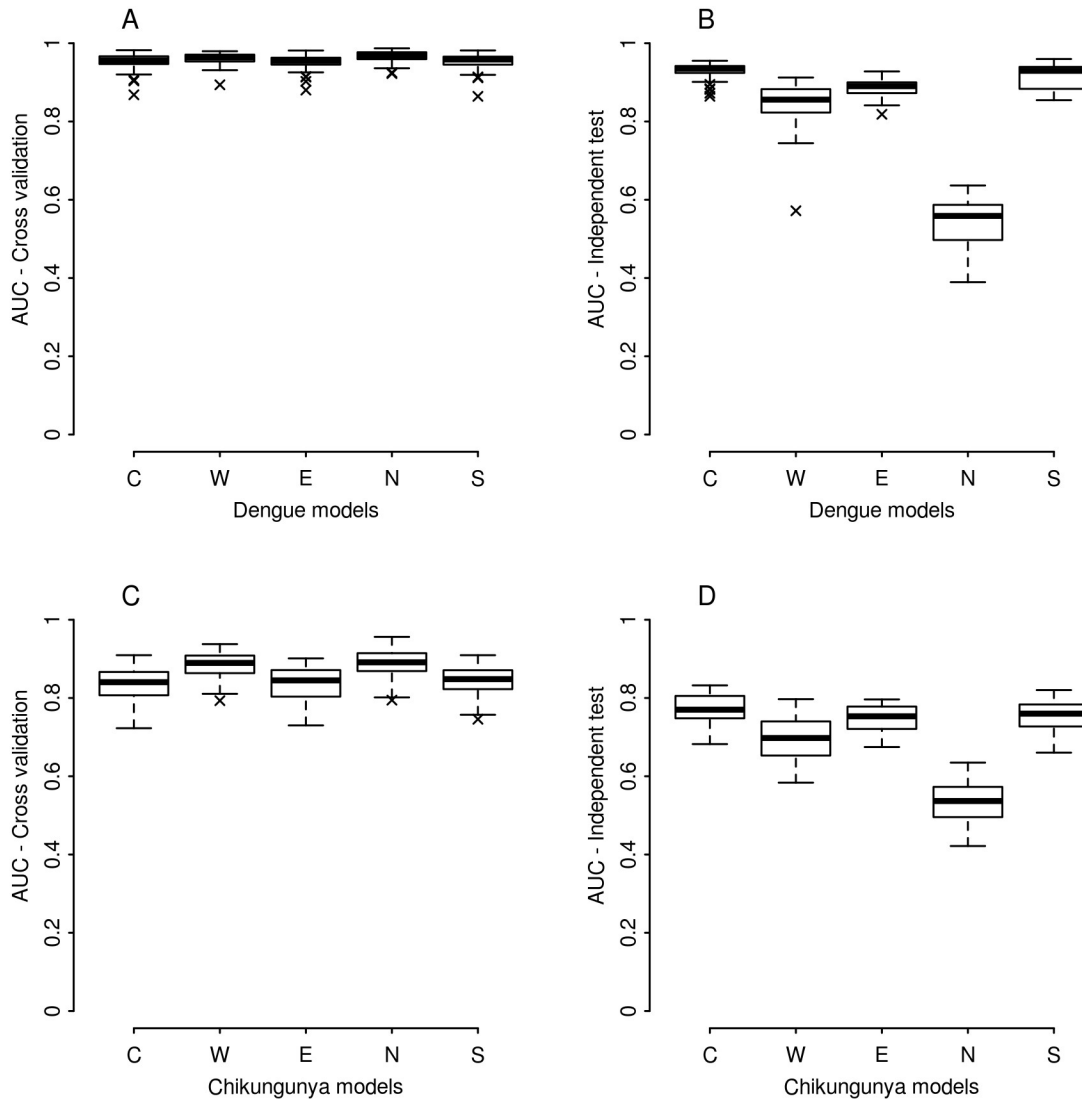
**Figure 2:** Analysis of the co-occurrences of dengue virus outbreaks within the Indian Ocean region from 1965 to 2009. A) The different relative importance of the considered covariates (CON, territory contiguity; CCO, common colonizer/part of the same country in the past; CLA, common language; DIS, geographical distance; CLI, climatic distance; DWE, GDP/wealth distance; DEG, density gravity index; MIG, migration), B) the fraction of the explained deviance by the 10-fold Cross Validation (CV) process (mean, red solid line, and standard error, black dashed line), and C) the number of positive co-occurrences in each 5-year period.



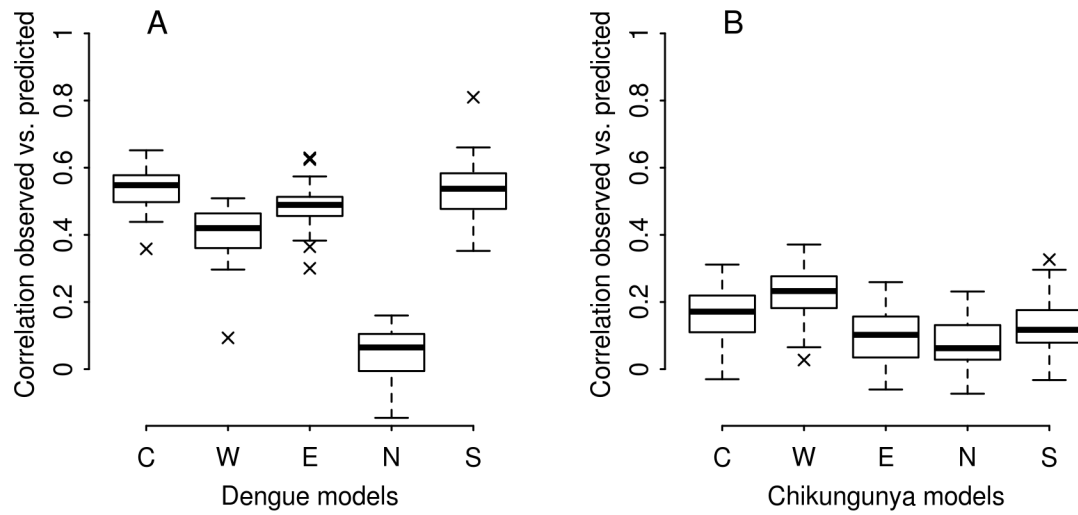
**Figure 3:** Analysis of the co-occurrences of chikungunya virus outbreaks within the Indian Ocean region from 1965 to 2009. A) The different relative importance of the considered covariates (CON, territory contiguity; CCO, common colonizer/part of the same country in the past; CLA, common language; DIS, geographical distance; CLI, climatic distance; DWE, GDP/wealth distance; DEG, density gravity index; MIG, migration), B) the fraction of the explained deviance by the 10-fold Cross Validation (CV) process (mean, red solid line, and standard error, black dashed line), and C) the number of positive co-occurrences in each 5-year period.



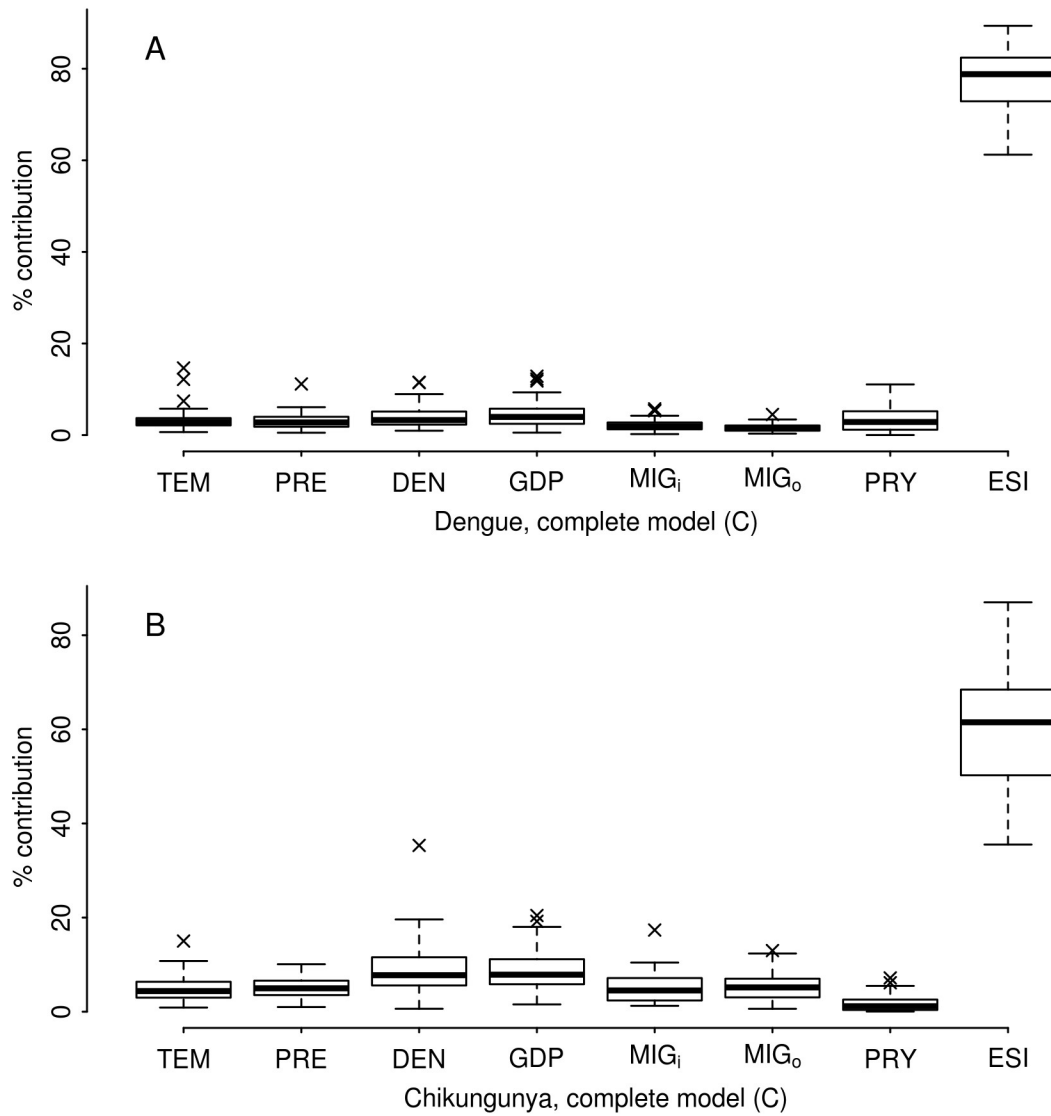
**Figure 4:** Analysis of the co-occurrences of chikungunya virus outbreaks within the Indian Ocean region in years from 2002 to 2009. A) The different relative importance of the considered covariates (CON, territory contiguity; CCO, common colonizer/part of the same country in the past; CLA, common language; DIS, geographical distance; CLI, climatic distance; DWE, GDP/wealth distance; DEG, density gravity index; MIG, migration), B) the fraction of the explained deviance by the 10-fold Cross Validation (CV) process (mean, red solid line, and standard error, black dashed line), and C) the number of positive co-occurrences in each 5-year period.



**Figure 5:** The Area under the receiver operating characteristic curve (AUC) in the analysed models: complete (C), no external force of infection (W), no within-country previous status (E), no past outbreak information (N), and simplified model (S). AUC in the cross-validation (CV, panels A and C), and in the independent nation-specific test (panels B and D) for the five dengue (panels A and B), and chikungunya (panels C and D) models. Boxplots represent the distribution of values for the 50 independent runs of the models, each time randomly selecting countries to be included in the training dataset.



621 **Figure 6:** Point biserial correlation between the predicted and observed outbreaks occurrences values  
622 for the five dengue (panel A), and chikungunya (panel B) models: complete (C), no external force of  
623 infection (W), no within-country previous status (E), no past outbreak information (N), and simplified  
624 model (S). Boxplots represent the distribution of values for the 50 independent runs of the models, each  
625 time randomly selecting countries to be included in the training dataset.



627 **Figure 7:** Covariates contributions for the complete dengue (A) and chikungunya (B) models. Boxplots  
 628 represent the distribution of values for the 50 independent runs of the models, each time randomly  
 629 selecting countries to be included in the training dataset.

630

Co-variate data				
Factor	Class	Units	Years	Source
Mean temperature (TEM)	Climatic	°C	1901-2012	<a href="https://datahelpdesk.worldbank.org/knowledge-base/articles/902061-climate-data-api">Climate Data API (https://datahelpdesk.worldbank.org/knowledge-base/articles/902061-climate-data-api)</a>
Mean precipitation (PRE)	Climatic	mm	1901-2012	<a href="https://datahelpdesk.worldbank.org/knowledge-base/articles/902061-climate-data-api">Climate Data API (https://datahelpdesk.worldbank.org/knowledge-base/articles/902061-climate-data-api)</a>
Population (POP)	Demographic	#individuals	1966-2014	<a href="http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators">Data WorldBank (http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators)</a>
Mean pop. Density (DEN)	Demographic	#ind./km <sup>2</sup>	1966-2014	<a href="http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators">Data WorldBank (http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators)</a>
Gross domestic product (GDP)	Economic	2005 USD	1960-2014	<a href="http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators">Data WorldBank (http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators)</a>
Migration (MIG)	Demographic	#individuals	1960-2009	Estimation in Abel 2013 and Abel 2014
Distance (DIS)	Geographical	km	-	<a href="http://www.cepii.fr/">CEPII (http://www.cepii.fr/)</a>
Contiguity (CON)	Geographical	binary	-	<a href="http://www.cepii.fr/">CEPII (http://www.cepii.fr/)</a>
Colonial/Language similarity (CCO/CLA)	Historic	binary	-	<a href="http://www.cepii.fr/">CEPII (http://www.cepii.fr/)</a>

631

632 **Table 1:** Description and source of the selected covariates data.